Developing Data Health & Maturity with Open-Source Technology

Frederic Randall

Data Manager & IT Specialist

Maryland New Directions

Glossary

**A/B Testing** – A user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing as used in the field of statistics.

**Application Programming Interface (API)** – API’s simplify software development and innovation by enabling applications to exchange data and functionality easily and securely.

**Agile Culture** – An environment that is underpinned by core values, behaviors, and practices, enabling all levels of an organization to better adapt to cultural, strategic, and other changes.

**Customer Relationship Management (CRM) Software** – A tool designed to build better relationships by providing a complete picture of all customer interactions.

**Dashboard** – A visual display of all your data, configurable to which data you want to see and whether you want to include charts or graphs to visualize numbers.

**Data Aggregation** – The compiling of information from multiple databases with intent to prepare combined datasets for data processing.

**Data Drift** – Unexpected and undocumented changes to data structure, semantics, and infrastructure.

**Data Fragmentation** – The state in which information storage space, main storage, or secondary storage, is used inefficiently, reducing capacity or performance and often both.

**Data Health** - The reflection of data’s ability to support decisions and objectives

**Data Literacy** – The ability to read, understand, create, and communicate data as information.

**Data Maturity** – The reflection of an organization’s capacity to perform varying degrees of data analysis.

**Descriptive Analytics** – The examination of data or content, usually manually performed, to answer the question “What Happened?”

**Diagnostic Analytics** – A form of advanced analytics that examines data or content to answer the question, “Why did it happen?”

**Email Analytics** – A method of tracking the way subscribers interact with your email campaigns.

**Fundraising and Donor Analytics** – The method used to describe the behavior of your donors and classify them into groups.

**Lean Methodology** – A way of optimizing the people, resources, effort, and energy of your organization toward creating value for the customer.

**Open-Source Software (OSS)** – Code that is designed to be publicly accessible for observation, modification, and distribution.

**Perishable Data** – Information that is subject to change and intended to reflect a current status.

**Predictive Analytics** – A form of advanced analytics that encompasses a variety of statistical techniques from data mining, predictive modelling, and machine learning that analyze current and historical facts to make predictions about the future.

**Prescriptive Analytics** – A form of advanced analytics which examines data or content to answer the question “What should be done?”

**Python** – An interpreted high-level general-purpose programming language.

**Relational Database** – A common type of database that stores data in tables, so it can be used in relation to other stored datasets.

**Relational Database Management System (RDBMS)** – A type of database management system that stores data in a row-based table structure which connects related data elements.

**Structured Query Language (SQL)** – Also known as sequel, a domain-specific language used in programming and designed for managing data held in a relational database management system.

**Web Analytics** – The measurement, collection, analysis, and reporting of web data to understand and optimize web usage.

Common Metrics in Advanced Analytics

**Donor Acquisition Rate** – Total of donor outreach costs divided by the number of acquired donors.

**Email Deliverability Rate** – The success rate an email marketer has in getting an email to a person’s correct email address.

**Conversion Marketing** – Marketing with the intention of increasing conversions from electronic commerce.

**Propensity to Give (P2G)**– P2G scores give an idea of how much a person can donate and how likely they are to do so at your organization.

**Social Return on Investment (SROI)**– Method for measuring values that are not traditionally reflected in financial statements, including social, economic, and environmental factors.

Open-Source Tools

Relational Database Management Systems

* PostgreSQL
* MariaDB
* CockroachDB
* MongoDB
* ClickHouse
* Neo4j

Platforms and Integrated Development Environments

* Anaconda
* Jupyter Notebooks
* Visual Studio Code
* Sublime Text
* Google Colaboratory

Python Libraries

* Pandas
* Matplotlib
* NumPy
* BeautifulSoup

Learning Resources

Communities & Free Online Resources

* Stack Overflow
* GitHub
* Kaggle
* Codewars
* HackerRank
* Kahn Academy
* Toptal
* Mozilla Developer Network
* *Automate The Boring Stuff with Python*
  + By Al Sweigart

Recommended Reading

* *Data Driven Nonprofits* by Steve Maclaughlin
* *Python Crash Course: A Hands-On, Project Based Introduction to Programming* by Eric Matthes
* *The Lean Startup* by Eric Ries
* *The Networked Nonprofit: Connecting with Social Media to Drive Change* by Allison Fine and Beth Kanter
* *Applied Data Science with Python and Jupyter* by Alex Galea
* *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* by Eric Siegel
* *How to Lie with Statistics* by Darrell Ruff
* *Naked Statistics: Stripping the Dread from the Data* by Charles Wheelan
* *Evidence and Narrative* by Edward Tufte

Other Sources

* *The Future of Nonprofit Fundraising and Engagement Survey Report* – EveryAction
* *Nonprofit Trends Report 2020* - Salesforce
* *Disempowered by Data: Nonprofits, Social Enterprises, and the Consequences of Data-Driven Work* by Chris Bopp, Ellie Harmon, Amy Voida
* *Win Big with Big Data: A Nonprofit’s Guide to Data Analytics* by Garrett Daniel and David Schlaseman
* *What is Open Source?*  - Red Hat